J. Montreuil, H. Schachter and J.F.G. Vliegenthart (Eds.), *Glycoproteins*

CHAPTER 2

Primary structure of glycoprotein glycans

JOHANNES F.G. VLIEGENTHART[1] and JEAN MONTREUIL[2]

[1]*Bijvoet Center for Biomolecular Research, Department of Bio-organic Chemistry, P.O. Box 80.075, 3508 TB Utrecht, the Netherlands*
[2]*Université des Sciences et Technologies de Lille, Laboratoire de Chimie Biologique (UMR no. 111 du CNRS), 59655 Villeneuve d'Ascq Cedex, France*

1.1. Introduction

For a long time, research on the primary structure of glycans was restricted to mucins, polysaccharides of connective tissue, serum glycoproteins and egg white glycoproteins. The reasons for such a peculiar choice of compounds are found in the inherent limitations of the analytical methods. Until the 1970s, large quantities of starting material were needed, due to the low sensitivity of the applied techniques. However, as mentioned in Chapter 1, the subsequent improvements of methods for isolation and structure determination led to the characterization of thousands of primary glycan structures. In this respect, it is relevant to note that the data bank CarbBank [1] contains at the moment over 38 000 records, of which 15 000 are unique structures. On the average hundreds of novel structures have been described every year since 1991. In addition, the discovery by biologists of the fundamental importance of glycoconjugates, has contributed to the spread of the research to many other species (lower animals, parasites, yeasts, fungi and bacteria). Most of the currently known glycan structures of glycoproteins are presented in the following chapters of this book.

1.2. Intact glycoproteins

The determination of the primary structure of glycoproteins comprises the analysis of the protein sequence, the identification of the glycosylation sites, the unravelling of the glycan structures and the determination of the microheterogeneity of the glycans at each glycosylation site. For these studies it is essential that adequate starting material is available and that the glycoprotein sample to be analysed contains a collection of glycoforms similar to that occurring in the natural environment. In practice this is a difficult problem because particular glycoforms may be (partly) lost or newly created due to degradation during the isolation and fractionation procedures. For soluble glycoproteins, the standard techniques for protein isolation and purification are first followed, as reviewed in [2]. A glycoprotein that is apparently homogeneous with respect

to the protein backbone, may nevertheless contain a collection of glycoforms [3]. Complete fractionation of all glycoforms is difficult to achieve, in particular when more than one glycosylation site is present and each of these sites exhibits microheterogeneity. Nevertheless, by lectin- or antibody-based affinity chromatography [2] at least partial fractionation can be obtained. High performance capillary electrophoresis [4] looks promising for realizing further separation of glycoforms. However, it can be envisaged that this technique also has inherent limitations.The isolation and purification of membrane bound glycoproteins give rise to specific problems that cannot be solved in a general way, and often detergents and/or chaotropic reagents have to be used. Glycoproteins that are membrane bound via glycosylphosphatidyl inositol (GPI) anchors are usually analysed after cleavage from the anchor. There are several ways to achieve this cleavage [5,6].

The detailed analysis of the primary structures of the glycans can hardly be carried out at the level of the intact glycoprotein. However, electrospray–mass spectrometry (ES-MS) has been shown to be a useful technique to establish post-translational modifications of proteins [7], even if the compounds have a molecular mass larger than 100 kDa. For various glycoproteins the ES-MS profile has been recorded and has yielded information about the glycoforms present, especially by coupling capillary electrophoresis as a separation technique with ES-MS as a detector [8]. For glycoproteins with a molecular mass up to 20 kDa, NMR spectroscopy can reveal many structural details [9–14]. Recently, the application of gradient-enhanced natural abundance ^{1}H-^{13}C HSQC and HSQC-TOCSY spectroscopy has been shown to be effective for the assignment of the NMR resonances of the carbohydrate chains of an intact glycoprotein [15]. By the former technique, the anomeric ^{1}H-^{13}C correlations can be derived, and the latter approach allows the deduction of correlations of the monosaccharide skeleton atoms.

1.3. Partial structures

To derive the primary structures of the entire collection of glycan chains attached to a protein, it is usually necessary to degrade the glycoprotein to partial structures like oligosaccharides, oligosaccharide–alditols or glycopeptides. For the last type of compounds it is important that the number of glycosylation sites comprised in the structure is known. There have been approaches described to obtain partial structures from N,O-glycoproteins (for a review see Ref. [16]).

However, two main problems have to be solved: (i) how to achieve a sequential, complete cleavage of N- and O-linked chains; and (ii) how to establish the positions in the protein chain of the glycosylation sites. The degradation to partial structures can be performed with chemical or enzymic methods. Hydrazinolysis is an example of a chemical method that is frequently applied to liberate N-linked chains as oligosaccharides [17,18]. Disadvantages of this methodology may be: (i) chemical modifications at the reducing end, which enhance the heterogeneity; (ii) partial cleavage of O-linked chains in N, O-glycoproteins; and (iii) removal of several essential non-carbohydrate substituents. Careful manipulation of the reaction conditions may reduce these compli-

cations and it has been claimed that it is possible to apply graded hydrozinolysis leading to sequential release of the O-linked chains followed by the N-linked chains. For each glycoprotein, the situation might be different, necessitating always a careful check of the effectiveness of the methodology. The release of O-linked chains, e.g. from O-glycoproteins in the form of oligosaccharide alditols can be realized with alkaline borohydride [19].

In the enzymic approach peptide-N^4-(*N*-acetyl-β-D-glucosaminyl) asparagine amidase (PNGase) is currently the most frequently used enzyme to liberate the N-chains as oligosaccharides, although endoglycosidases are also applied. Depending on the glycoprotein PNGase F or A treatment is used. After PNGase digestion of N,O-glycoproteins, the remaining O-glycoprotein can be isolated and then subjected to alkaline borohydride degradation [20]. It is essential to check the completeness of the enzymic cleavage carefully e.g. by SDS-PAGE with silver staining, or with the aid of a glycan detection kit, a Con-A test in a microtiter plate [21] or a test with another lectin in an analogous way. Often the incubation conditions like pH, the addition of chaotropic or reducing agents and for large glycoproteins a preceding, partial proteolysis have to be optimized for each glycoprotein in order to ensure a complete release of the N-linked chains.

In all cases described above the released glycans are obtained as complex mixtures of oligosaccharides or oligosaccharide alditols. This may create severe fractionation problems, that require multiple chromatographic steps. To obtain pure compounds reduction of the oligosaccharides to alditols is sometimes carried out to simplify the complexity of the mixture.

In some instances endoglycosidases are used to split off oligosaccharides from N-linked chains. The specificity of these enzymes can be rather strict, thereby limiting their applicability [22].

Various proteases are in use to degrade N-glycoproteins to glycopeptides in order to obtain pure glycopeptides for the determination of the glycosylation sites and of the site-specific (micro)heterogeneity [23]. The amino acid sequence, the glycan structure and the specificity of the protease control the size and type of the glycopeptides formed. Often incomplete and aspecific proteolytic cleavages give rise to complex mixtures of peptides and glycopeptides. This may afford redundant results, because one glycosylation site can now be represented by more than one glycopeptide. In O-, and N,O-glycoproteins often multiple O-glycosylation sites occur. These sites are frequently clustered, thereby impeding the preparation of glycopeptides containing only a single O-glycosylation site. In heavily O-glycosylated proteins the assignment of a particular glycan to a specific amino acid can be difficult.

The fractionation of partial structures has been improved enormously by the introduction of: (i) better column materials for gel permeation chromatography to achieve size fractionation [24]; (ii) lectin-affinity chromatography [25]; (iii) HPLC on anion exchange materials [26]; (iv) high pH anion exchange chromatography (HPAEC) [27], in combination with pulsed amperometric (PAD) or radiometric detection (this method is sensitive and can be applied to small amounts of material); (v) high performance capillary electrophoresis [28].

1.4. Structure determination

To define the structure of glycans completely, the following parameters have to be determined:

(i) Type and number of the constituent monosaccharides, including absolute configuration, ring size and anomeric configuration.

(ii) Sequence of the monosaccharides, including the positions of the glycosidic linkages.

(iii) Type, number and location of non-carbohydrate substituents.

A variety of methods are currently used to establish these parameters depending on the type of problems to be solved, and on the availability of material and techniques. For analysis of the monosaccharide composition, the samples are often subjected to methanolysis followed by re-*N*-acetylation, trimethylsilylation and GLC [29]. Alternatively, acid hydrolysis can be carried out followed by HPAEC-PAD [27]. Linkage analysis is carried out on partially methylated monosaccharide alditols. They are obtained by permethylation of the sample, e.g., in DMSO with a suitable base and CH_3I, followed by hydrolysis, e.g., in 4M trifluoro acetic acid, acetylation and analysis by GLC-MS [30,31]. Exoglycosidases are employed to gain information on the non-reducing-end monosaccharides with regard to identity and absolute and anomeric configuration. Sequential enzymic degradation with exoglycosidases can provide some insight into the structure [32]. However, these exoglycosidases are not very specific as to ring size, linkage position and branching point. Furthermore, the presence of non-carbohydrate substituents might inhibit their action. It is a prerequisite that detailed knowledge exists on the specificity of the exo-enzymes. Endoglycosidases with known substrate specificity can provide additional information [16,22]. For oligosaccharides and glycopeptides advanced mass spectrometric techniques like FAB-MS; ES-MS; MALD-MS; MALD-TOF-MS; MS-MS are suitable to obtain structural information as to branching pattern, number and length of branches and sequence. The mass spectrometric analysis of different derivatives and/or chemically modified compounds may furnish complementary information. A significant advantage of mass spectrometry is that only low amounts of material are required [7].

High resolution ^{1}H-NMR spectroscopy at 500 or at 600 MHz has become a powerful method for the identification of N- as well as O-type carbohydrate chains, due to the introduction of the structural-reporter-group concept [33–38]. In D_2O, the greater part of the skeleton protons of the constituting monosaccharides resonate in a narrow window between δ 3.5 and δ 3.9, thereby giving rise to a bulk signal. The signals outside this region are the structural-reporter-group signals that contain the essential information for translation into primary structures. They comprise: (i) anomeric protons; (ii) protons attached to carbon atoms close to the substitution position in a monosaccharide constituent (so-called glycosylation shifts); (iii) protons at deoxy-carbon atoms; (iv) *N*-acetyl and *N*-glycolyl methylene protons; (v) protons shifted as a result of the presence of non-carbohydrate substituents. In addition to 1D-NMR spectra, various types of 2D-NMR spectra like COSY, HOHAHA, NOESY, HMQC and HMBC are used to arrive at unambiguous assignments. The large number of ^{1}H-NMR data that are available for glycoprotein derived carbohydrate chains, has

enabled the construction of a NMR database that is connected to the complex carbohydrate structure database [39]. Owing to the enormous variety of glycoprotein derived glycans that have been identified now, it has become feasible to develop oligosaccharide-finger-printing methods, facilitating reliable and fast batch control of, e.g., biotechnologically produced glycoproteins. These methods include different separation techniques like high pH anion-exchange chromatography or capillary electrophoresis [40–45]. Another possibility is offered by PAGE of oligosaccharide derivatives obtained after reductive amination with 8-aminonaphthalene 1,3,6-trisulfonic acid [46]. However, in the cases where novel structures are involved, it is essential to have pertinent structural data obtained via more than a single technique, in order to arrive at unambiguous conclusions.

The presence of non-carbohydrate substituents may give rise to specific analytical problems, starting with the isolation of the glycoprotein. It is essential that during the isolation and purification procedures the authentic non-carbohydrate substituents remain present and are not split off and do not migrate to other positions. The substituents identified so far in glycoproteins can be methyl, acetyl, lactyl, glycolyl, aminoethyl phosphonate, sulfate or phosphate groups. For the determination of the identity and location of these substituents various methods are in use, but the most detailed information is obtained from NMR spectroscopy and mass spectrometry (cf. Refs. [20,47]).

2.1. Monosaccharide constituents

For a long time, the 'classical' monosaccharides constituting the glycans were D-galactose, D-mannose, D-glucose, L-fucose, D-xylose, L-arabinofuranose, *N*-acetyl-D-glucosamine, *N*-acetyl-D-galactosamine and some sialic acids. However, refinement of the analytical methods has led to the characterization of new monosaccharides which were first considered as rare sugars, but which now appear to be more common than previously thought (for review, see Ref. [48]), e.g., 2-keto-3-deoxy-nonulosonic acid (Kdn), one of the 35 known sialic acids [49], first discovered in fish eggs [50] and recently characterized in batracian eggs [51].

The monosaccharides which are likely to be found in glycoprotein glycans, other than the above-mentioned 'classical' ones, are listed in Table I.

2.2. The glycan protein linkages

Glycans are conjugated to peptide chains by two types of primary covalent linkages: N-glycosyl and O-glycosyl linkages as listed in Table II.

Until recently, the only N-glycosyl bond characterized in glycoproteins regardless of whether they originated from animals, plants, microorganisms or viruses, was the *N*-acetylglucosaminyl–asparagine bond formally resulting from linking a β-*N*-acetylglucosamine residue to the amido group of L-asparagine, as discovered by Neuberger et al. [75] and almost simultaneously by others [76,77]. However, during recent years,

TABLE I

Monosaccharides recently identified in glycoprotein glycans [4]

Monosaccharides	Sources	Refs.
2-Acetamido-4-amino-2,4,6-trideoxyglucose	*Clostridium symbiosum*	[52]
6-Deoxylaltrose	Salmonid fish eggs	[53]
3-Deoxy-D-*glycero* -*galacto* -nonulosonic acid (Kdn)	Salmonid fish eggs	[50]
2,3-Diacetamido-2,3-dideoxy-mannuronic acid	*Bacillus stearothermophilus*	[52]
2-*O*-Methylfucose	Nematodes	[54]
Galactofuranose	Bacteria	[55,56]
	Trypanosoma	[57]
3-*O*-Methylgalactose	Yeasts	[58]
	Snail	[59]
4-*O*-Methylgalactose	Nematodes	[54]
6-*O*-Methylgalactose	Algae	[60]
Galactose-3-sulfate	Thyroglobulin	[61,62]
	Mucins of cystic fibrosis	[63]
N-Acetylgalactosamine-4-sulfate	Pituitary hormones	[64]
	Tamm Horsfall glycoprotein	[65]
	Urokinase	[66]
3-*O*-Methylglucose	*Methanothermus fervidus*	[67]
3-*O*-Methyl-*N*-acetylglucosamine	*Clostridium thermocellum*	[56]
N-Acetylglucosamine-6-sulfate	Thyroglobulin	[61,62]
Gulose	Algae	[63]
3-*O*-Methylmannose	Snail	[69]
Mannose-4-sulfate	Ovalbumin	[70]
Mannose-6-sulfate	Ovalbumin	[70]
	Slime mold	[71]
Mannose-6-phosphate Me	Slime mold	[72]
N-Acetylmannosamine	*Clostridium symbosium*	[52]
4,8-Anhydro-neuraminic acid	Edible bird's nest	[73]
8-*O*-Methyl 9-*O*-acetyl-*N*-glycolyl-neuraminic acid	Starfish	[74]
8-*O*-Methyl-7,9-*O* -diacetyl-*N*-glycolylneuraminic acid	Starfish	[74]

linkages between asparagine and other sugars such as glucose, *N*-acetylgalactosamine and L-rhamnose [48] have been characterized in bacterial glycoproteins.

Unambiguous proof for the existence in glycoproteins of O-glycosidic linkages of glycans to serine and threonine was provided in the early 1960s. The evidence was based on the alkali-lability of the sugar-β-hydroxyamino acid linkage which is split by a β-elimination reaction.

The *N*-acetylgalactosamine-serine/threonine bond was first demonstrated in the mucins [78–81] and is widely distributed in nature in the so-called mucin-type glycoproteins. The O-glycosidic carbohydrate-serine linkage was discovered by Helen Muir in 1958 [82] in chondroitin-4-sulfate and the involvement of xylose was demonstrated

TABLE II

Carbohydrate-peptide linkages in glycoproteins [a]

Type of glycoprotein	Amino acid	Monosaccharide
N-glycosylproteins		
	Asn	β-GlcNAc
	Asn	β-GalNAc
	Asn	α/β-Glc
	Asn	L-Rha
O-glycosylproteins		
Mucin type	Ser/Thr	α-GalNAc
Intracellular type	Ser/Thr	β-GlcNAc
Yeast type	Ser/Thr	*x*-Man
Rat tissues	Ser/Thr	α-L-Fuc
Proteoglycan type	Ser	β-Xyl
Worm collagen type	Ser	α-Gal
Factor IX type	Ser	β-Glc
Collagen type	OH-Lys	β-Gal
Extensin type	OH-Pro	β-L-Ara*f*
Algal type	OH-Pro	β-Gal
Glycogenin type	Tyr	α-Glc
S-layer-type[b]	Tyr	β-Glc

(a) For detailed historical review, see Refs. [48] and [105].
(b) In crystalline surface layer of *Clostridium thermohydrosulfuricum*.

in 1964 by Lindahl and Rodén in heparin [83] and in chondroitin-4-sulfate [84]. At present, the xylosyl–serine linkage is considered to be characteristic of animal proteoglycans. Other glycan–amino acid bonds are: (i) galactosyl (β1–5) hydroxylysine in guinea-pig-skin tropocollagen [85,86]; (ii) L-arabinofuranosyl (β1–4) hydroxyproline first recognized by Lamport et al. in cell walls from higher plants and in tomato extensin [87,88], by Neuberger et al. in potato lectin [89] and by others in a very wide variety of higher- and lower-plant cell tissues.

Two novel bonds have been added to the list of O-linked amino acids: the glucosyl (α1–O) tyrosine linkage in glycogenin [90] and *N*-acetylglucosaminyl (β1–O) serine in intracellular glycoproteins (nuclear pore, chromatin proteins, transcription factors and cytoplasmic inclusions); interestingly no other sugars are attached to the *N*-acetylglucosamine residue (for review, see Ref. [91]). In this regard, two additional novel bonds have been characterized. The first one originates from the non-enzymic condensation of a monosaccharide like glucose with the ε-amino group of lysine residues leading to the formation of the imino group of a Schiff's base. This reaction is called *glycation*. The second one concerns a new class of widely distributed membrane components of eukaryotic cells, parasites in particular: the glycosylphosphatidyl inositol anchor (GPI) [92,93]. This kind of attachment has been termed *glypiation*.

Recently a new type of linkage between a carbohydrate and a protein has been found, involving C-glycosylation of a specific tryptophan residue in human RNase U_S [94]. The origin of this linkage has still to be traced.

2.3. *The inner-core and antenna concept*

On the basis of the glycan primary structures of 14 N-glycosylproteins analyzed in 1974 and using the observation that all of them presented the common structural pattern of the mannotriose-di-*N,N'*-acetyl chitobiose (Fig. 1) linked to an asparagine residue, Montreuil [97] proposed in 1975 to designate this common and non-specific invariant fraction (*inv* fraction) *the core* and the variable oligosaccharide structure (*var* fraction) planted on the inner-core, the *antennae*. The latter are mobile and support the biological activity of the glycans.

The concept of core was further extended to various O-glycoproteins and the most classical ones are presented in Fig. 2. It is interesting to mention the hypothesis of Jett and Jamieson [96] according to which the *N*-acetylglucosamine asparagine linkage is the most primitive one and the other amino acids involved in the glycan–protein

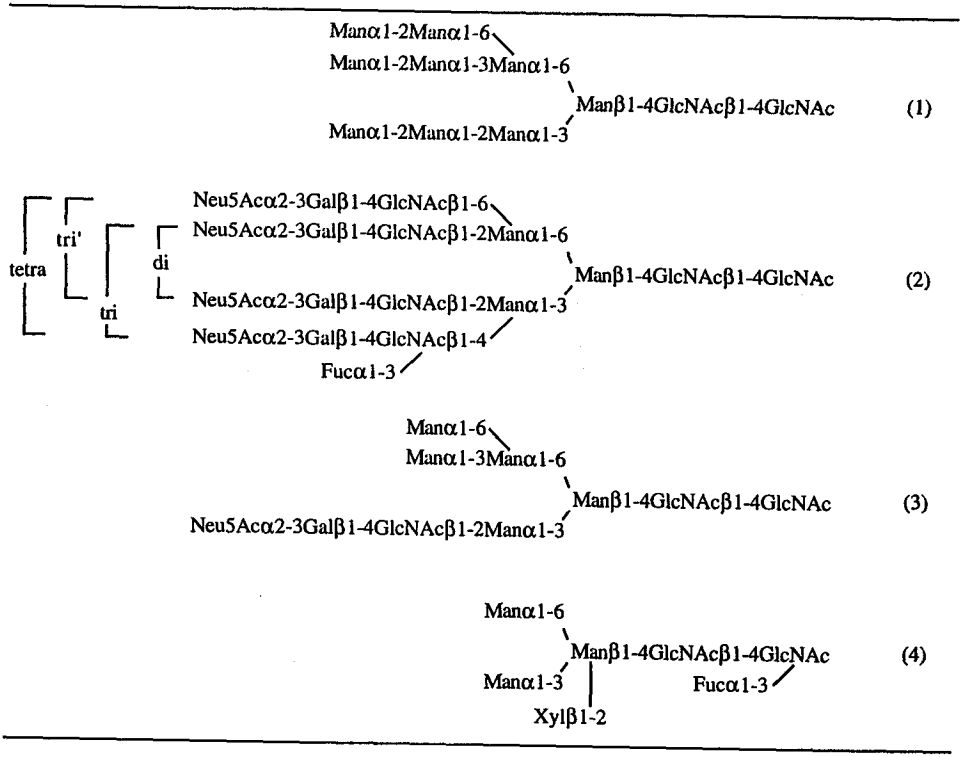


Fig. 1. Examples of the four types of N-linked carbohydrate chains of glycoproteins. 1. Oligomannose type; 2. *N*-acetyllactosamine type (complex type); 3. Hybrid type; 4. Xylose-containing type.

Type	Structure	Type	Structure
core 1	Galβ1-3GalNAc	core 4	GlcNAcβ1-6 GlcNAcβ1-3GalNAc
core 2	GlcNAcβ1-6 Galβ1-3GalNAc	core 5	GalNAcα1-3GalNAc
core 3	GlcNAcβ1-3GalNAc	core 6	GlcNAcβ1-6 GalNAc

Fig. 2. Core structures of mucin type O-linked carbohydrate chains of glycoproteins.

linkages, except hydroxyproline could be derived from asparagine by a single-mutation of the original codons AAU and AAC into AGU and AGC for serine, ACU and ACC for threonine, AAA and AAG for lysine.

2.4. Classification and nomenclature of glycans and glycoproteins

For a long time the nomenclature and definition of glycoproteins remained chaotic because little was known before the 60s about the composition and the primary structure of glycans and about the carbohydrate–protein linkage (for review, see Refs. [97,105]). However, little by little these parameters became better defined leading to a series of definitions that are at present known as the recommendations of the Joint Commission on Biochemical Nomenclature of the International Union of Pure and Applied Chemistry (IUPAC) and the International Union of Biochemistry and Molecular Biology (IUBMB). These recommendations have been published in various periodicals. The main definitions are the following.

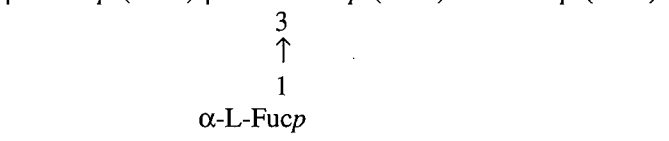
A glycoprotein is a compound containing carbohydrate (or glycan) covalently linked to protein. The carbohydrate may be in the form of a monosaccharide, disaccharide(s), oligosaccharide(s), polysaccharide(s), or their derivatives (e.g. sulfo- or phospho-substituted). One, a few, or many carbohydrate units may be present. Proteoglycans are a subclass of glycoproteins in which the carbohydrate units are polysaccharides that contain amino sugars. Such polysaccharides are also known as glycosaminoglycans. A glycopeptide is a compound consisting of carbohydrate linked to an oligopeptide composed of L- and/or D-amino acids. A glyco-amino-acid is a saccharide attached to a single amino acid by any kind of covalent bond. A glycosyl-amino-acid is a compound consisting of saccharide linked through a glycosyl linkage (O-, N-, or S-) to an amino acid. (The hyphens are needed to avoid implying that the carbohydrate is necessarily linked to the amino group.)

In N, O-glycoproteins the prefix N- is used for the N-glycosyl linkage to asparagine. N-linked oligosaccharides are divided into two major classes: the *N*-acetyllactosamine

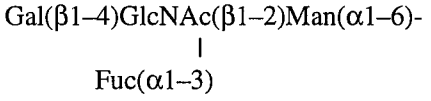
type containing *N*-acetyl-D-glucosamine, D-mannose, D-galactose, L-fucose and sialic acid; and the oligomannose type containing *N*-acetyl-D-glucosamine and a variable number of D-mannose residues. Structures containing both oligomannose- and *N*-acetyllactosamine-type oligosaccharides are designated as hybrid type. The prefix O- is used for O-glycosyl linkage to serine, threonine, hydroxylysine or bydroxyproline.

Two types of carbohydrate–peptide linkage in the same protein or peptide chain may be indicated by a combination of the prefixes. For the condensed representation of sugar chains, the non-reducing terminus of the carbohydrate chain should always be on the left-hand end. Current practice allows the use of either an extended form (a) or a condensed form (b), which allows structures to be shown in one line as well as in two or more, and in which the longest chain should always be the main chain:

(a) Extended form



(b) Condensed form in two lines



or condensed form in one line

Gal(β1–4)[Fuc(α1–3)]GlcNAc(β1–2)-

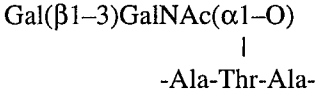
To describe large structures, the short form has been proposed:

(c) Short form

Galβ4(Fucα3)GlcNAcβ2Manα6-

The short form is obtained: (i) by omitting locants of anomeric carbon atoms, (ii) omitting the parentheses around the specification of linkage, and (iii) omitting hyphens.

Similarly a glycopeptide sequence, represented in the condensed form in two lines as



or in the condensed form in one line as

-Ala-[Gal(β1–3)GalNAc(α 1–O)]Thr-Ala-

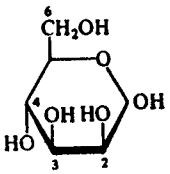
may be written in the short form in two lines as

Galβ3GalNAcα —
-Ala-Thr-Ala

N-linked oligosaccharides contain a common pentasaccharide core as follows:

Manα 6 \
> Manβ4GlcNAcβ4GlcNAc
Manα 3 /

For the sake of uniformity, the location of substitution should be written as above in accordance with Haworth's representation of the pyranose structure of monosaccharides.



In the condensed system of symbols for sugar residues (as used in this book) the common configuration and ring size (usually pyranose) are implied in the symbol. Thus, Gal denotes D-galactopyranose; Man, D-mannopyranose; Fuc, L-fucopyranose; GlcNAc, 2-acetamido-2-deoxy-D-glucopyranose or *N*-acetyl-D-glucosamine; Neu5Ac (which may be abbreviated to NeuAc), *N*-acetylneuraminic acid. The symbol Sia stands for sialic acid, a general term that can also be used when the exact structure is unknown.

Whenever the configuration or ring size is found to differ from the common one it must be indicated by using the appropriate symbols for the extended system. The configuration of amino acids is L unless otherwise noted. Although symbols such as Gal and Man are useful in representing oligosaccharide structures they should not be used in the text to represent monosaccharides.

2.5. Microheterogeneity of glycans

In addition to genetically determined variants expressed as variations in their polypeptide chains, almost all glycoproteins reveal another form of polymorphism associated with their glycan moieties. This type of diversity is termed microheterogeneity or peripheral heterogeneity because it involves the number and positions of the most external monosaccharides in the glycan moieties. These variants, recently called *glycoforms* [3], were first characterized in α_1-acid glycoprotein from human serum by Karl Schmid [99] by electrophoresis. The microheterogeneity of α_1 acid glycoprotein was found to be due to the occurrence of di-, tri-, and tetra-antennary glycans of the *N*-acetyllactosamine type at the 5 glycosylation sites [100]. The glycans can be fuco-

sylated [101] and sialylated at different levels. This feature is wide spread and has been observed for natural as well as for recombinant DNA glycoproteins [cf. 47]. Microheterogeneity can also be present in the O-linked glycans, of the mucin type, as shown for example for the pig zona pellucida glycoproteins [102].

The existence of microheterogeneity gives rise to a lot of interesting questions regarding the origin of this phenomenon and about its relevance for the biological functioning of the glycoforms that can be distinguished. For gaining further insight, it is at least essential to have an exact knowledge of the naturally occurring microheterogeneity at each glycosylation site. In case recombinant DNA glycoproteins are produced for therapeutic purposes it is required that the glycan chains of the produced glycoproteins are compatible with the immune system of the patient. Preferably, the glycosylation pattern should be identical or close to that of the natural glycoprotein.

2.6. Evolution

Our present knowledge of the primary structures of numerous glycans raises interesting problems from a comparative biochemical point of view and confirms some of the concepts developed by Montreuil [95,103,104]. Glycan structures from a given class or order of living organisms have common oligosaccharides and are not randomly constructed at all. On the contrary they are subject to rules based on a conservative evolution of glycosyltransferases and on the high specificity and subcellular localization of these enzymes.

3. Concluding remarks

In the area of structural glycobiology the development is going very fast. The availability of sophisticated methods for the isolation and characterization of glycoproteins and glycoprotein-derived glycans has paved the way for the study of the complete three-dimensional structure in solution of low molecular mass glycoproteins by means of a combination of NMR techniques and molecular dynamics simulations. It is in particular the insight into the spatial structures that is needed to make significant progress in the understanding of structure–function relations for carbohydrate chains. Another fascinating area that will become feasible for study is the interaction between glycoproteins and receptors. Although those systems represent usually high-molecular mass complexes, they should be studied in order to analyse in which way the carbohydrate chains are involved in the interaction processes. In situations wherein both the carbohydrate and the protein moieties participate in the interaction with receptors, it is relevant to learn if the moieties are recognized one after another or that the a 'domain type' of recognition occurs wherein both are involved at the same time. The dynamic aspects of recognition in relation to the response to the interaction form challenging goals of further work.

Abbreviations

Con A	concanavaline A
COSY	correlated spectroscopy
ES-MS	electrospray–mass spectrometry
FAB-MS	fast atom bombardment–mass spectrometry
GLC-MS	gas liquid chromatography–mass spectrometry
GPI	glycosylphosphatidyl inositol
HMBC	heteronuclear multiple bond correlation
HMQC	heteronuclear multiple quantum correlated spectroscopy
HOHAHA	homonuclear Hartmann–Hahn
HPAEC	high pH anion exchange chromatography
HPLC	high performance liquid chromatography
HSQC	^{1}H-detected heteronuclear single quantum coherence spectroscopy
MALD-MS	matrix assisted laser desorption–mass spectrometry
MALD-TDF-MS	matrix assisted laser desorption–time of flight–mass spectrometry
NMR	nuclear magnetic resonance
NOESY	nuclear Overhauser enhancement spectroscopy
PAD	pulsed amperometric detection
PNGase	peptide-N^4-(*N*-acetyl(β-D-glucosaminyl) asparagine amidase
S-layer	surface layer
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
TOCSY	total correlated spectroscopy

References

1. Doubet, S., Bock, K., Smith, D., Darvill, A. and Albersheim, P. (1989) TIBS 14, 475–477.
2. Carlsson, S.L.R. (1993) In: M. Fukuda and A. Kobata (Eds.), Glycobiology: A Practical Approach. Oxford University Press, Oxford, UK, pp. 1–26.
3. Rademacher, T.W., Parekh, R.B. and Dwek, R.A. (1988) Annu. Rev. Biochem. 57, 785–838.
4. Rudd, P.M., Scragg, I.G., Coghill, E.C. and Dwek, R.A. (1992) Glycoconjugate J. 9, 86–91.
5. Ferguson, M.A.J. (1993) In: M. Fukuda and A. Kobata (Eds.), Glycobiology: A Practical Approach. Oxford University Press, Oxford, UK, pp. 349–383.
6. Menon, A.K. (1994) Methods in Enzymology 230, 418–442.
7. Dell, A., Reason, A.J., Khoo, K.H., Parrico, M., McDowell, R.A. and Morris, H. (1994) Methods in Enzymology 230, 108–132.
8. Kelly, J.F., Locke, S.J. and Thibault, P. (1993) Discovery 2 (2), 1–6.
9. Dill, K., Berman, E. and Pavia, A.A. (1985) Adv. Carbohydr. Chem. Biochem. 43, 1–49.
10. Brockbank, R.L. and Vogel, H.J. (1990) Biochemistry 29, 5574–5583.
11. Joao, H.C., Scragg, I.G. and Dwek, R.A. (1992) FEBS Lett. 307, 343–346.
12. Wyss, D.F., Withka, J.M., Knoppers, M.H., Sterne, K.A., Recny, M.A. and Wagner, G. (1993) Biochemistry 32, 10995–11006.
13. Withka, J.M., Wyss, D.F., Wagner, G., Arulanandam, A.R.N., Reinherz, E.L. and Recny, M.A. (1993) Structure 1, 69–81.

14. Fletcher, M.C., Harrison, R.A., Lachmann, P.J. and Neuhaus, D. (1994) Structure 2, 185–199.
15. De Beer, T., Van Zuylen, C.W.E.M., Hård, K., Boelens, R., Kaptein, R., Kamerling, J.P. and Vliegenthart, J.F.G. (1994) FEBS Lett. 348, 1–6.
16. Dwek, R.A., Edge, Chr.J., Harvey, D.J., Wormald, M.R. and Parekh, R.B. (1993) Annu. Rev. Biochem. 62, 65–100.
17. Kobata, A. and Endo, T. (1993) In: M. Fukuda and A. Kobata (Eds.), Glycobiology, A Practical Approach. Oxford University Press, Oxford, UK, pp. 79–102.
18. Patel, T.P. and Parekh, R.B. (1994) Methods in Enzymology 230, 57–66.
19. Piller, F. and Piller, V. (1993) In: M. Fukuda and A. Kobata (Eds.), Glycobiology, A Practical Approach. Oxford University Press, Oxford UK, pp. 291–328.
20. Vliegenthart, J.F.G., Hård, K., De Waard, P. and Kamerling, J.P. (1991) In: H.S. Conradt (Ed.), Protein Glycosylation: Cellular, Biotechnological and Analytical Aspects. VCH Publishers, Inc., New York, NY, USA.
21. Langeveld, J.P.M., Noelken, M.E., Hård, K., Todd, P., Vliegenthart, J.F.G., Rouse, J. and Hudson, B.G. (1991) J. Biol. Chem. 266, 2622–2631.
22. Tarentino, A.L. and Plummer, T.H. (1994) Methods in Enzymology 230, 44–57.
23. Lee, Y.C. and Rice, K.G. (1993) In: M. Fukuda and A. Kobata (Eds.), Glycobiology, A Practical Approach. Oxford University Press, Oxford, UK, pp. 127-163.
24. Kobata, A. (1994) Methods in Enzymology 230, 200–208.
25. Kobata, A. and Yamashita, K. (1993) In: M. Fukuda and A. Kobata (Eds.), Glycobiology, A Practical Approach. Oxford University Press, Oxford, UK, pp. 103–125.
26. Baenziger, J.U. (1994) Methods in Enzymology 230, 237–249.
27. Hardy, M.R. and Townsend, R.R. (1994) Methods in Enzymology 230, 208–225.
28. Linhardt, R.J. (1994) Methods in Enzymology 230, 265–280.
29. Kamerling, J.P. and Vliegenthart, J.F.G. (1989) In: A.M. Lawson (Ed.), Clinical Biochemistry; Principles, Methods, Applications. Vol. 1, Mass Spectrometry. Walter de Gruyter & Co, Berlin, Germany, pp. 175–263.
30. Cummings, R.D. (1993) In: M. Fukuda and A. Kobata (Eds.), Glycobiology, A Practical Approach. Oxford University Press, Oxford, UK, pp. 243–289.
31. Geyer, R. and Geyer, H. (1994) Methods in Enzymology 230, 86–108.
32. Jacob, G.S. and Scudder, P. (1994) Methods in Enzymology 230, 280–299.
33. Vliegenthart, J.F.G., Van Halbeek, H. and Dorland, L. (1980) In: A. Varmavuoni (Ed.), 27th Int. Congress, Pure and Appl. Chem. Pergamon Press, Elmsford, NY, USA, pp. 253–262.
34. Vliegenthart, J.F.G., Van Halbeek, H. and Dorland, L. (1981) Pure Appl. Chem. 53, 45–77.
35. Vliegenthart, J.F.G., Dorland, L. and Van Halbeek, H. (1983) Adv. Carbohydr. Chem. Biochem. 41, 209–374.
36. Kamerling, J.P. and Vliegenthart, J.F.G. (1992) Biological Magnetic Resonance 10, 1–194.
37. Hård, K. and Vliegenthart, J.F.G. (1993) In: M. Fukuda and A. Kobata (Eds.) Glycobiology, A Practical Approach. Oxford University Press, Oxford, UK, pp. 223–242.
38. Van Halbeek, H. (1994) Methods in Enzymology, 230, 132–168.
39. Van Kuik, J.A., Hård, K. and Vliegenthart, J.F.G. (1992) Carbohydr. Res. 235, 53–68.
40. Spelmann, M. (1990) Anal. Chem. 62, 1714–1722.
41. Taverna, M., Baillet, A., Biou, D., Schlütter, M., Werner, R., Fevrier, D. (1992) Electrophoresis 13, 359–366.
42. Suzuki, S., Kakehi, K. and Honda, S. (1992) Anal. Biochem. 205, 227–236.
43. Hermentin, P., Witzel, R., Doenges, R., Bauer, R., Haupt, H., Patel, T., Parekh, R.B. and Brazel, D. (1992) Anal. Biochem. 206, 419–429.
44. Hermentin, P., Doenges, R., Witzel, R., Hokke, C.H., Vliegenthart, J.F.G., Kamerling, J.P., Conradt, H.S., Nimtz, M. and Brazel, D. (1994) Anal. Biochem. 221, 29–41.

45. Rice, K.G., Takahashi, N., Namiki, Y., Tran, A.D., Lisi, P.J. and Y.C. Lee (1992) Anal. Biochem. 206, 278–287.
46. Jackson, P. (1994) Methods in Enzymology 230, 250–265.
47. Vliegenthart, J.F.G. (1994) In: K. Bock and H. Clausen (Ed.), Alfred Benzon Symposium 36, Complex Carbohydrates in Drug Research, Structural and Functional Aspects. Munksgaard, Copenhagen, Denmark, pp. 30–44.
48. Lis, H. and Sharon, N. (1993) Eur. J. Biochem. 218, 1–27.
49. Schauer, R. (1991) Glycobiology 1, 449–452.
50. Kanamori, A., Inoue, S., Iwasaki, M., Kitajima, K., Kawai, G., Yokoyama, S. and Inoue, Y. (1990) J. Biol. Chem. 265, 21811–21819.
51. Strecker, G., Wieruszeski, J.M., Michalski, J.C., Alonso, C., Leroy, Y., Boilly, B. and Montreuil, J. (1992) Eur. J. Biochem. 207, 995–1002.
52. Messner, P. and Sleytr, V.B. (1991) Glycobiology 1, 545–551.
53. Iwasaki, M., Inoue, S. and Inoue, Y. (1987) Eur. J. Biochem. 168, 185–192.
54. Khoo, K.H., Maizels, R.M., Page, A.P., Taylor, G.W., Rendell, N.B. and Dell, A. (1991) Glycobiology 1, 163–171.
55. Gerwig, G.J., Kamerling, J.P., Vliegenthart, J.F.G., Morag, E., Lamed, R. and Bayer, E.A. (1992) Eur. J. Biochem. 205, 799–808.
56. Gerwig, G.J., Kamerling, J.P., Vliegenthart, J.F.G., Morag (Morgenstern), E., Lamed, R. and Bayer, E.A. (1991) Eur. J. Biochem. 196, 115–122.
57. Moraes, C.T., Bosch, M. and Parodi, A.J. (1988) Biochemistry 27, 1543–1549.
58. Takanayagi, T., Kushida, K., Idonuma, K. and Ajisaka, K. (1992) Glycoconjugate J. 9, 229–234.
59. Van Kuik, J.A., Sybesma, R.P., Kamerling, J.P., Vliegenthart, J.F.G. and Wood, E.J. (1986) Eur. J. Biochem. 160, 621–625.
60. Schlipfenbacher, R., Wenzl, S., Lottspeich, F. and Sumper, M. (1986) FEBS Lett. 209, 57–
61. Spiro, R.G. and Bhoyroo, V.D. (1988) J. Biol. Chem. 263, 14351–14358.
62. De Waard, P., Koorevaar, A., Kamerling, J.P. and Vliegenthart, J.F.G., (1991), J. Biol. Chem. 266, 4237–4243.
63. Lamblin, G., Rahmoune, H., Wieruszeski, J.M., Lhermitte, M., Strecker, G. and Roussel, P. (1991) Biochem. J. 275, 199–206.
64. Baenziger, J.U. and Green, E.D. (1988) Biochim. Biophys. Acta 947, 287–306.
65. Hård, K., Van Zadelhoff, G., Moonen, P., Kamerling, J.P. and Vliegenthart, J.F.G. (1992) Eur. J. Biochem. 209, 895–915.
66. Bergwerff, A.A., Van Oostrum, J., Kamerling, J.P. and Vliegenthart, J.F.G. (1995) Eur. J. Biochem. in press.
67. Hartmann, E. and Koning, H. (1989) Arch. Microbiol. 151, 274–278.
68. Mengele, R. and Sumper, M. (1992) FEBS Lett. 298, 14–16.
69. Van Kuik, J.A., Sybesma, R.P., Kamerling, J.P., Vliegenthart, J.F.G. and Wood, E.J. (1987) Eur. J. Biochem. 169, 399–411.
70. Yamashita, K., Ueda, I. and Kobata, A. (1983) J. Biol. Chem. 258, 14144–14147.
71. Freeze, H.H. (1985) Arch. Biochem. Biophys. 243, 580–693.
72. Gabel, C.A., Costello, C.E. Reinhold, V.N., Kurz, L. and Kornfeld, S. (1984) J. Biol. Chem. 259, 13762–13769.
73. Pozzgay, V., Jennings, H.J. and Kasper, D.L. (1987) Eur. J. Biochem. 162, 445–450.
74. Bergwerff, A.A., Hulleman, S.H.D., Kamerling, J.P., Vliegenthart, J.F.G., Shaw, L., Reuter, G. and Schauer, R. (1992) Biochimie 74, 25–38.
75. Johansen, P.G., Marshall, R.D. and Neuberger, A. (1961) Biochem. J. 78, 518–527.
76. Nuenke, R.H. and Cunningham, L.W. (1961) J. Biol. Chem. 236, 2451–2460.

77. Yamashina, I. and Makino, M. (1962) J. Biochem. (Tokyo) 51, 359–364.
78. Hashimoto, Y. and Pigman, W. (1962) Ann. N.Y. Acad. Sci. 93, 541.
79. Tanaka, K., Bertolini, M. and Pigman, N. (1964) Biochem. Biophys. Res. Commun. 16, 404–409.
80. Anderson, B., Seno, H., Sampson, P., Riley, J.G., Hoffman, P. and Meyer, K. (1964) J. Biol. Chem. 239, PC 2716.
81. Bhavanandan, V.P., Buddecke, E., Carubelli, P. and Gottschalk, A. (1964) Biochem. Biophys. Res. Commun. 16, 353–357.
82. Muir, H. (1958) Biochem. J. 69, 195–204.
83. Lindahl, U and Roden, L. (1964) Biochem. Biophys. Res. Commun. 17, 254–259.
84. Roden, L. and Lindahl, U. (1965) Fed. Proc. 24, 606.
85. Butler, W.T. and Cunningham, L.W. (1955) J. Biol. Chem. 240, PC 3449.
86. Butler, W.T. and Cunningham, L.W. (1966) J. Biol. Chem. 240, 3882–3888.
87. Lamport, D.T.A. (1967) Nature 216, 1322–1324.
88. Lamport, D.T.A. and Miller, D.H. (1971) Plant Physiol. 48, 454–456.
89. Allen, A.K., Desai, N.N., Neuberger, A. and Creeth, J.M. (1978) Biochem. J. 171, 665–674.
90. Smythe, C., Caudwell, F.G., Ferguson, M. and Cohen, P. (1988) EMBO J. 7, 2681–2686.
91. Hart, G.W., Haltiwanger, R.S., Holt, G.D. and Kelly, W.G. (1989) Annu. Rev. Biochem. 58, 841–874.
92. Low., M.G., Ferguson, M.A.J., Futerman, A.H. and Silman, I. (1986) Trends Biochem. Sci. 11, 212–215.
93. Eglund, P.T. (1993) Annu. Rev. Biochem. 62, 121–138.
94. Hofsteenge, J., Müller, D.R., De Beer, T., Löffler, A., Richter, W.J. and Vliegenthart, J.F.G. (1994) Biochemistry 33, 13524–13530.
95. Montreuil, J. (1975) Pure Appl. Chem. 42, 431–477.
96. Jett, M. and Jamieson, G.A. (1971) Carbohydr. Res. 18, 466–468.
97. Kisailus, E.C. and Allen, H.J. (1992) In: H.J. Allen and E. Kisailus (Eds.), Nomenclature of Monosaccharides, Oligosaccharides, and Glycoconjugates: Composition, Structure, and Function. Marcel Dekker Inc., New York, USA, pp. 13–32.
98. Nomenclature of glycoproteins, glycopeptides and peptidoglycans:
(a) Eur. J. Biochem. 159, 1–6 (1986), correction in 185, 485 (1989).
(b) Glycoconjugate J. 3, 123–134 (1986).
(c) J. Biol. Chem. 262, 13–18 (1967).
(d) Pure Appl. Chem. 60, 1389–1384 (1988).
(e) Spec. Period. Rep. Amino Acids Pept., Proteins, 21, 329–334 (1994).
(f) Biochemical Nomenclature and Related Documents, A Compendium, 2nd ed. 1992. Portland Press, London, UK, pp. 84–89.
99. Schmid, K., Binette, J.P., Kamiyama, S., Pfister, V. and Takahashi (1962) Biochemistry 1, 959–956.
100. Fournet, B., Strecker, G., Montreuil, J., Vliegenthart, J.F.G., Schmid, K. and Binette, J.P. (1978) Biochemistry 17, 5206–5214.
101. Van Halbeek, H., Dorland, L., Vliegenthart, J.F.G., Montreuil, J., Fournet, B. and Schmid, K. (1981) J. Biol. Chem. 256, 5588–5590.
102. Hokke, C.H., Damm, J.B.L., Penninkhof, B., Aitken, R.J., Kamerling, J.P., and Vliegenthart, J.F.G. (1994) Eur. J. Biochem. 221, 491–512.
103. Montreuil, J., Spik, G. and Chasson, A (1962) C.R. Acad. Sci. Paris 255, 3493–3494.
104. Montreuil, J., Adam-Chasson, A. and Spik G. (1965) Bull. Soc. Chim. Biol. 47, 1867–1880.
105. Montreuil, J. (1982) In: A. Neuberger and L.L.M. van Deenen (Eds.), Glycoproteins, Comprehensive Biochemistry 19B, Part II. Elsevier, Amsterdam, pp. 1–188.